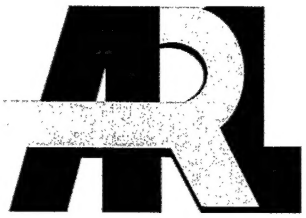
ARMY RESEARCH LABORATORY



Stereo Camera Re-calibration and the Impact of Pixel Location Uncertainty

William F. Oberle

ARL-TR-2979 **MAY 2003**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturers' or trade names does not constitute an official endorsement or approval of the use thereof.

DESTRUCTION NOTICE—Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

ARL-TR-2979 May 2003

Stereo Camera Re-calibration and the Impact of Pixel Location Uncertainty

William F. Oberle
Weapons & Materials Research Directorate

INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT DATE	3. DATES COVERED (From - To)
May 2003	Final	

4. TITLE AND SUBTITLE

Stereo Camera Re-calibration and the Impact of Pixel Location Uncertainty

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Oberle, W.F. (ARL)

5d. PROJECT NUMBER
622618.H03

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
Weapons & Materials Research Directorate
Aberdeen Proving Ground, MD 21005-5066

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TR-2979

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report presents an algorithm for the re-calibration of a stereo camera system mounted on an autonomous vehicle. The algorithm employs feature-based matching in combination with an iterative approach to estimating the fundamental matrix with the RANdom SAmple Consensus (RANSAC) paradigm followed by a constrained nonlinear estimate. Simulations indicate that stereo camera re-calibration is possible with the algorithm, and the accuracy of the process depends on the amount of uncertainty in the pixel location of corresponding left and right image points. Future work will be directed at improving the accuracy of determining corresponding points.

15. SUBJECT TERMS

essential matrix fundamental matrix RANSAC stereo camera calibration stereopsis

16. SECURITY CLASSIFICATION OF			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			William F. Oberle
Unclassified	Unclassified	Unclassified	UL	35	19b. TELEPHONE NUMBER *(Include area code)* 410-278-4362

INTENTIONALLY LEFT BLANK

Contents

List of Figures

List of Tables

Acknowledgments

The author would like to thank Gary A. Haas and Dr. Mary Anne Fields of the U.S. Army Research Laboratory for their time and effort in reviewing this report. Their comments, observations, and recommendations are greatly appreciated.

INTENTIONALLY LEFT BLANK

1. Introduction

Stereo vision or stereopsis is a scheme for depth perception that permits depth or distance information of a scene to be determined from two or more images of the scene taken from different viewpoints. When incorporated as a component of the sensor suite on autonomous or unmanned ground vehicles (UGV) in a military application, this methodology offers the potential to provide information related to vehicle navigation, reconnaissance, surveillance, and target identification. However, the density and accuracy of the information obtained depend from a computational standpoint on the fidelity of the solution to the stereo *correspondence* and *reconstruction* problems.[1] A flow diagram of the stereopsis paradigm illustrating the relationship of the *correspondence* and *reconstruction* problems together with required input is provided in Figure 1. In order to render the *correspondence* problem more tractable, generally the left and right images are rectified so that corresponding image points are situated on the same horizontal scan line, reducing the search space for potential matches from two dimensional (2-D) to one dimensional. At a minimum, the rectification procedure requires knowledge of the external (extrinsic) camera parameters, which are defined as the rotation matrix and translation vector from one camera's coordinate system or reference frame to the reference frame of the second camera. In addition to the external camera parameters, unambiguous (absolute three-dimensional [3-D] coordinates, no scale factor or projective transformation required) solutions of the *reconstruction* problem require knowledge of the intrinsic parameters for both cameras.[2] Thus, the quality of the information obtained from stereopsis depends on the accuracy of the estimates for the extrinsic and intrinsic camera parameters.

Obtaining the extrinsic and intrinsic camera parameters is termed "camera calibration". The classical approach to camera calibration is to analyze the stereo images of a surveyed calibration pattern in different poses. A substantial amount of work has been performed in this area, and the reader is referred to Faugeras (1993), Xu and Zhang (1996), Trucco and Verri (1998), Gennery (2001), or Oberle and Haas (2002) and the references therein for a detailed explanation of this calibration process. However, if the stereo vision system is being used on a moving platform such as a UGV traversing off-road terrain, the camera parameters will probably begin to deviate from those calculated during the initial calibration procedure. Therefore, to maintain the accuracy of the stereopsis information, the stereo system should be re-calibrated on a regular

[1] "*The correspondence problem*: Which parts of the left and right images are projections of the same scene element? *The reconstruction problem*: Given a number of corresponding parts of the left and right images, and possibly information on the geometry of the stereo system, what can we cay about the 3-D location and structure of observed objects?" (Trucco & Verri 1998, p. 140).

[2] The intrinsic parameters for each camera are the ratios of the focal length to the horizontal and vertical pixel length, the coordinates of the principal point of the camera, and the angle between retinal axes. For most cameras this angle is 90 degrees (Zhang, Deriche, Faugeras, & Luong 1994).

schedule. However, such re-calibrations with the classical calibration approach may not be feasible, e.g., during operations in which the stereo system is required to operate for long periods of time without "hands-on" maintenance. Yet, even in these cases, re-calibration is still possible if certain conditions are met.

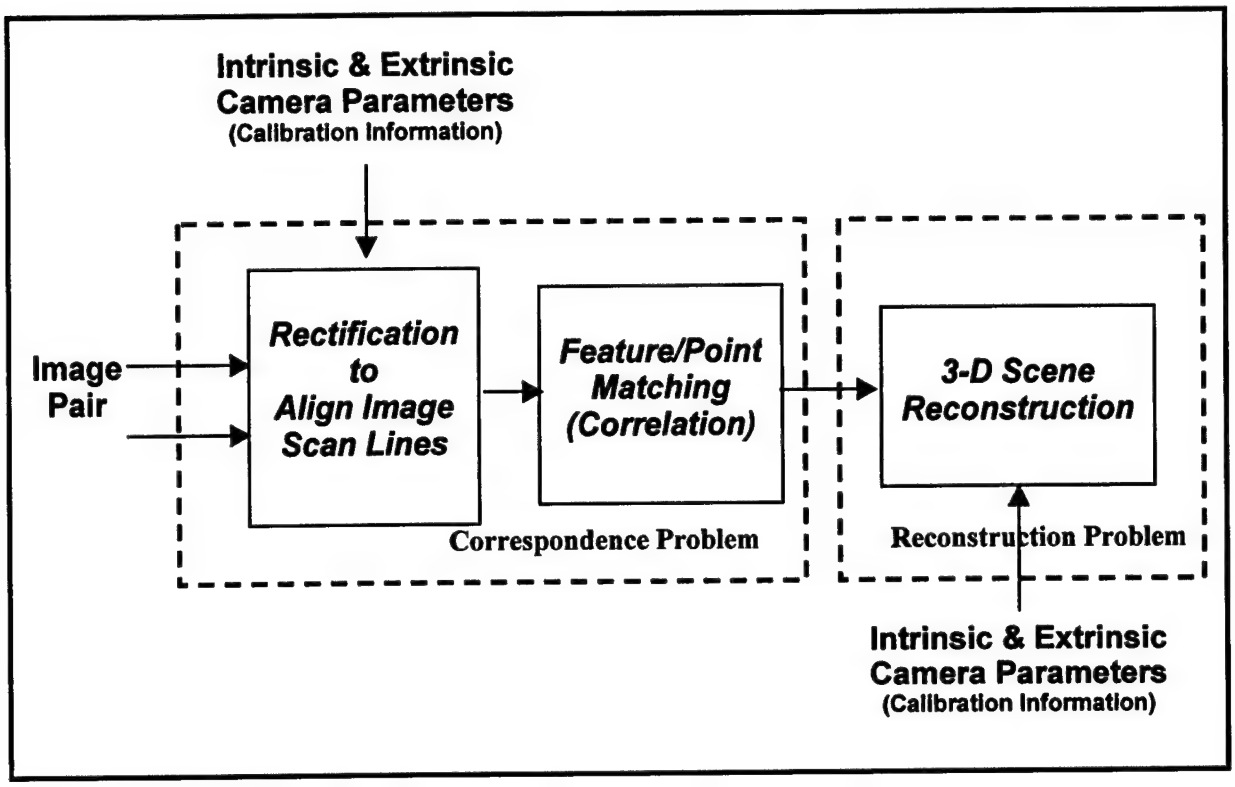


Figure 1. Stereopsis paradigm flow diagram.

A re-calibration of the extrinsic camera parameters is possible if all the following conditions are satisfied:

1. The intrinsic parameters of both cameras are known.
2. At least eight corresponding point matches between the left and right images are identified.
3. The distance between two distinct points in the scene is known.

Furthermore, if the intrinsic parameters for both cameras are assumed equal, then both the extrinsic and intrinsic camera parameters can be re-calibrated[3] (Torr 2002). For this work, it is assumed that changes in the intrinsic camera parameters from the original calibration (obtained

[3] This is often referred to as self-calibration.

with the classical approach) are sufficiently small. Thus, the intrinsic camera parameters are assumed to be known.[4]

The objectives of this report are twofold. First, to present an algorithm for stereo system camera re-calibration, assuming the intrinsic camera parameters are known, which could be incorporated on a UGV or other platform. The second objective is to assess the impact that uncertainty in the pixel location (of the corresponding point matches between the left and right images) has on the algorithm's results. This objective is motivated by the fact that whatever technique (e.g., corner matching) is used to determine the required corresponding point matches is subject to pixel location errors. Besides pixel location error, false matches or outliers may also be present. To compensate for this type of matching error, a robust technique in computing the fundamental matrix, which contains the extrinsic parameter information, is used. The organization of the remainder of this report is as follows. In Section 2, the proposed algorithm is presented and discussed. Implementation details concerning the calculation of the fundamental matrix are provided in Section 3. Section 4 contains the analysis that assesses pixel location error. Finally, a summary of this work is provided in Section 5.

2. Proposed Re-construction Algorithm

The re-construction algorithm now presented is drawn from ideas and approaches contained in the work by Trucco and Verri (1998), Torr (2002), and Loy (2002). A flow diagram for the algorithm is provided in Figure 2.

As shown in Figure 2, the algorithm consists of five major steps:

Step 1: Obtain a set of at least eight corresponding image points. Input is the stereo image pair.

Step 2: Estimate the fundamental matrix via the set of corresponding image points determined in Step 1.

Step 3: Compute the essential matrix from the fundamental matrix of Step 2 under the assumption that the intrinsic parameters of the cameras are known. Use singular value decomposition to modify the essential matrix to enforce the rank = 2 and equal singular values constraint.

Step 4: Determine candidate rotations and translations consistent with the modified essential matrix of Step 3. Use back projection of a single image point to determine the appropriate rotation-translation pair. This translation is estimated *modulo* a scale factor.

[4] If in practice this assumption proves to be inaccurate, the approach to re-calibration discussed in subsequent sections will have to be modified.

Step 5: Eliminate the unknown scale factor, based on the distance between two points in the scene.

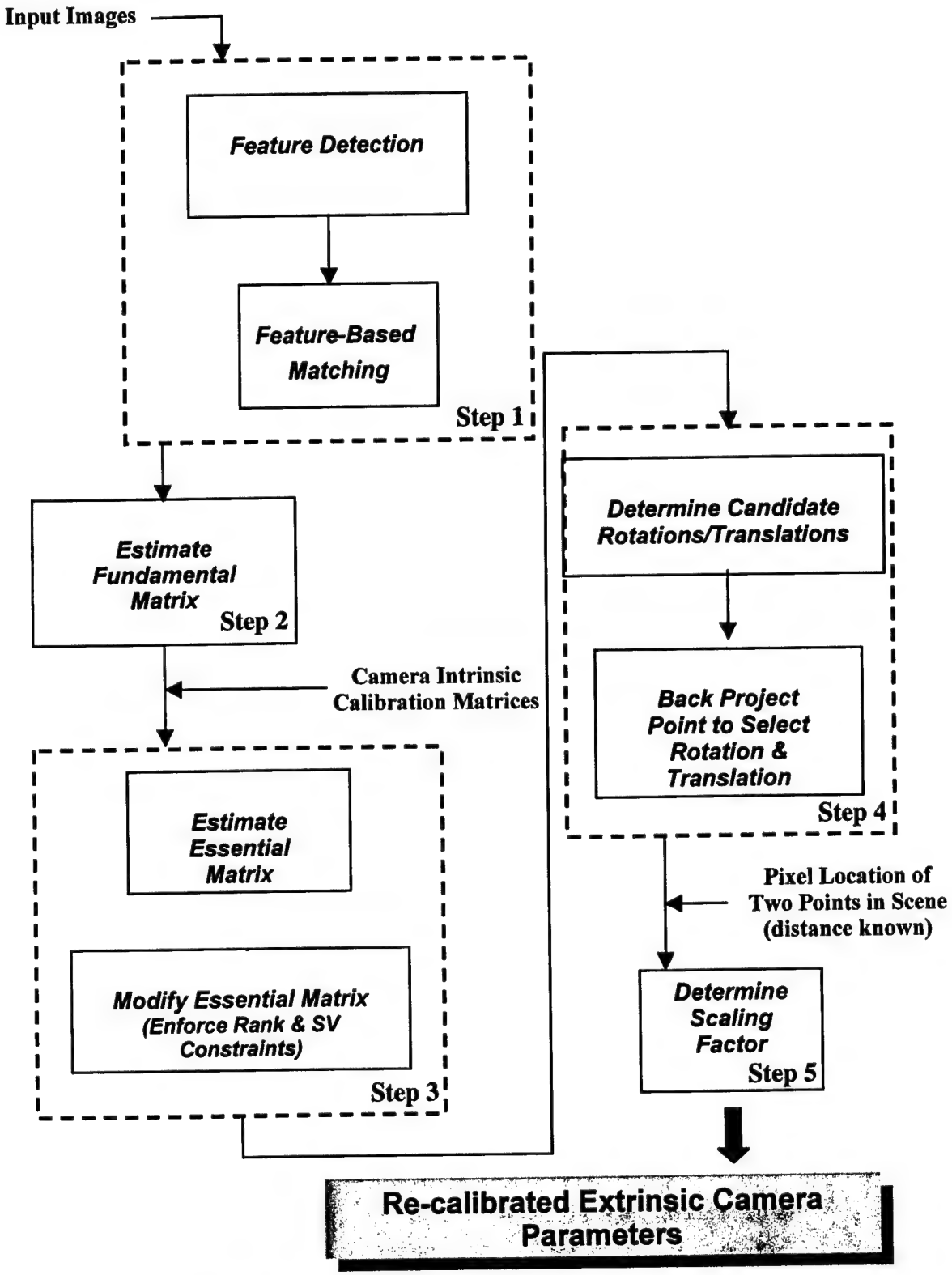


Figure 2. Flow diagram for re-calibration algorithm.

Step 1 is essentially the stereo correspondence problem without any of the constraints associated with structure models (e.g., epipolar geometry) since the extrinsic parameters of the stereo camera pair are not known. Two common approaches to address the correspondence problem are intensity-based matching and feature-based matching. To be efficient, intensity-based matching requires the images to be rectified. This is not possible when the extrinsic camera parameters are unknown as in this situation. Thus, for the algorithm, a feature-based matching approach is recommended. Corners are chosen as the feature to match because of their robustness relative to changes in perspective, ease of detection, and low computational cost to identify compared to other possible features. Also, the narrow baseline of stereo systems mounted on UGVs should mitigate difficulties with occlusion associated with corner detection. Other possible features that could be used include edges, line segments, curve segments, closed curves or regions. Although all the possible choices for the feature offer advantages and disadvantages, it is felt that selecting corners as the feature provides the best compromise among the competing features. If the use of corners as the matching feature proves to be inadequate, then edges are recommended as an additional feature to be incorporated. Finally, based on a literature review, the Harris corner detector (Harris & Stephens, 1988) appears to be a robust and widely used corner detector. For example, Torr (2002) incorporates this corner detector in his structure and motion toolkit.

Once a list of features from each image is compiled, the problem of matching the corresponding features (feature-based matching) in each image must be addressed. Unfortunately, the matching procedure is an ill-posed problem in the sense that there is no match for some features (e.g., because of occlusion); others may be matched with several corresponding features, while still others will be mismatched. In addition, even correctly matched points may suffer from pixel location error because of measurement inaccuracies. Clearly, features for which no match is achieved will not be included in a list of matched features. The difficulty of multiple matches can be avoided by the inclusion of only the best match (based on whatever matching criteria are being used). Thus, there are two sources of error in the matched features: mismatched features and correctly matched features with incorrect pixel locations. Mismatched features and correctly matched features with "large"[5] pixel location errors are termed "outliers"[6]. Unfortunately, the severity/number of the outliers affects the accuracy of the fundamental matrix calculation. Smith, Sinclair, Cipolla, and Wood (1998) demonstrated that the use of sub-pixel correlation windows together with the sum of squared differences (in place of cross correlation) as the

---

5 The magnitude of large and small when we are referring to pixel location error is determined by the user. It is controlled by the selection of a number of values used in the calculations, e.g., correlation window size and threshold value.

6 This is somewhat a misuse of the term outlier. Generally, a data point, even a correctly identified/measure point, is identified as an outlier if it does not satisfy a criterion associated with the mathematical model being used to describe the data. In this case, all matched points satisfy the mathematical model associated with the matching. It is the fact that these points should not be consistent with the mathematical model used to estimate the fundamental matrix that results in their being labeled outliers.

measure of similarity significantly reduces the percent of outliers by as much as 50%. It is recommended that their approach be incorporated in the matching process. The impact of correctly matched features with "small" pixel location errors or noise is addressed in Section 4.

In Step 2, the fundamental matrix for the stereo system is determined with the corresponding matches from Step 1. Although many methods are available to estimate the fundamental matrix (Zhang, 1996; Torr, 2002), the possibility of a large number of outliers resulting from the matching process dictates that a method that eliminates or adequately accounts for the outliers be incorporated into the procedure. Such methods are termed "robust estimators". As discussed in Lacey, Pinitkarn, and Thacker (2000) and Torr (2002), the RANdom SAmple Consensus (RANSAC) paradigm is an example of an effective robust algorithm for the stereopsis application of interest. Torr (2002, p 46) uses a modification of the basic RANSAC algorithm, MAPSAC (Maximum *A Posteriori* SAmple Consensus) that he claims "yields a modest to hefty benefit to all robust estimations with absolutely no additional computational burden". The use of the MAPSAC algorithm is recommended for use in the computation of the fundamental matrix. Theoretically, this should provide an accurate estimation for the fundamental matrix. However, in practice, this does not appear to be the case. The basic difficulty is that algorithms (seven point, eight point, least squared, iterative, etc.) for estimating the fundamental matrix (one of which must be used in RANSAC/MAPSAC; Torr uses the seven-point algorithm) appear to be susceptible to numerical instabilities associated with ill-conditioned matrix operations and/or local minima. Despite the pre-conditioning of the data (Hartley 1997; Trucco & Verri 1998; Torr 2002), calculations by the author using MAPSAC as implemented in Torr (2002) resulted in successive calculations with the same input data producing substantially different estimates for the fundamental matrix even with synthetic input data subject only to numerical noise associated with the numerical errors resulting from the computer calculations in generating the data. MATLAB[7] (2001) is used for all calculations.

To address this problem, an approach suggested by Torr (2002, p 14) is implemented: "...I suggest using a robust estimator like (sic) MAPSAC to get a first pass at $\mathbf{F}$[8] and then perform a constrained nonlinear estimation afterwards to optimize..." Unfortunately, similar results to the MAPSAC-only calculations (different estimates with the same input data) were obtained. However, running this calculation (i.e., MAPSAC followed by a second method) a fixed number of times with the same input data and selecting the result associated with the "minimal error" produces satisfactory results[9] for the data sets tested. Details of the implementation and error metric used are discussed in Section 3.

---

[7] MATLAB® is a registered trademark of The MathWorks.

[8] $\mathbf{F}$ is the fundamental matrix.

[9] Results are considered satisfactory if synthetic data are used: (1) the yaw, roll, and pitch angles for the computed rotation matrix are within 0.01 milliradian of the corresponding angles for the true rotation matrix, and (2) the angle between the computed and true translation vector is within 0.01 milliradian.

Finally, it is noted that the assumptions used to construct the fundamental matrix are sufficient only to define the fundamental matrix *modulo* a scale factor (Xu & Zhang 1996). This is reflected in the translation vector of Step 4 also being defined *modulo* a scale factor. This completes Step 2.

Step 3 consists of estimating the essential matrix. Although the procedure for obtaining the essential matrix from the fundamental matrix is straightforward, it is not unique in the sense that the order of several of the matrix operations can be reversed. Thus, specific details of the recommended procedure are provided.

First, the order of the transformation between the left and right camera coordinate systems is specified. Let $p_l = (X_l, Y_l, Z_l)$ and $p_r = (X_r, Y_r, Z_r)$ denote the 3-D coordinates of the same world point in the coordinate system of the left and right cameras, respectively. For this work, the transformation between p_l and p_r is defined by

$$p_r = \mathbf{R}(p_l - \mathbf{T}), \tag{1}$$

in which $\mathbf{R}$ represents the rotation matrix and $\mathbf{T}$ the translation vector between the left and right camera coordinate systems. Determining $\mathbf{R}$ and $\mathbf{T}$ is the goal of the re-calibration.

Although the re-calibration process estimates the extrinsic parameters that define a transformation in 3-D space, input data are in terms of a 2-D coordinate system associated with the camera image with length in pixels. The relationship between this 2-D pixel coordinate system, several intermediate coordinate systems, and the 3-D coordinate system together with the necessary information required to transform between coordinate systems is shown in Figure 3. More specific details concerning the transformation between the various coordinate systems is provided throughout this section.

In Step 2, a principal assumption (Torr 2002) for computing the fundamental matrix is that for any 3-D world point, if p_l' is the image in homogeneous pixel coordinates in the left image and p_r' is the corresponding point in the right image, then

$$p_r'^{T} \mathbf{F} p_l' = 0. \tag{2}$$

$\mathbf{F}$ is the fundamental matrix. The standard embedding of $\Re^2$ into 2-D projective space is $(x,y) \mapsto (x,y,1)$. However, Torr (2002) suggests using $(x,y) \mapsto (x,y,\zeta)$ with ζ chosen to obtain the best numerical conditioning. The effect of a different choice for ζ on the estimation of the fundamental matrix is discussed in Section 3. The standard embedding is used in the derivation.

Assuming that the transformation between 3-D points in the coordinate systems of the left and right cameras is given by Equation (1), then the essential matrix, $\mathbf{E}$, satisfies the equation (Xu & Zhang 1996; Trucco & Verri 1998)

$$\bar{p}_r^{T} \mathbf{E} \bar{p}_l = 0. \tag{3}$$

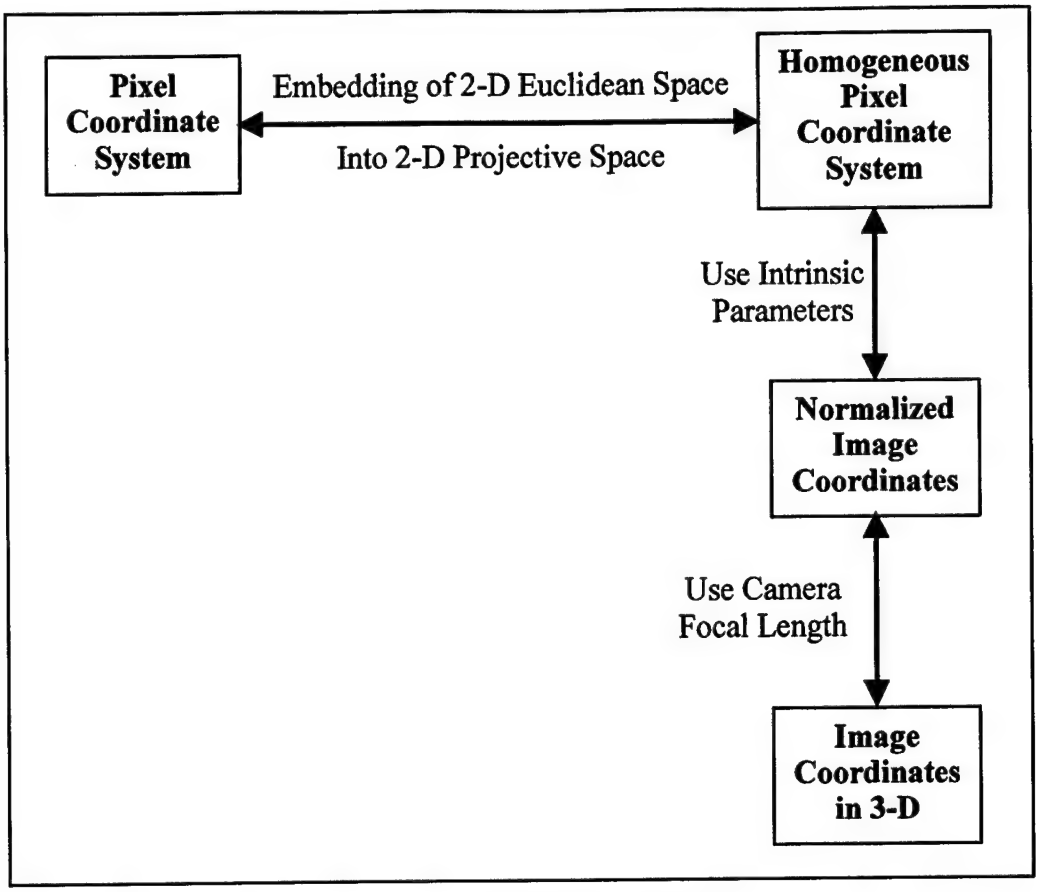


Figure 3. Coordinate systems used in re-calibration process.

Here, $\bar{p}_r$ and $\bar{p}_l$ represent normalized image points[10] defined by

$$\bar{p}_r = \frac{p_r}{Z_r} \text{ and } \bar{p}_l = \frac{p_l}{Z_l}. \tag{4}$$

Therefore, a relation between $\mathbf{F}$ and $\mathbf{E}$ can be established from Equations (2) and (3) since the normalized image points and the homogeneous pixel coordinates for each camera are related by the camera calibration or intrinsic matrix denoted by $\mathbf{K}$. An assumption for this work is that the calibration matrix has been determined for each of the cameras used in the stereo pair. The calibration matrix for a camera is defined (Xu & Zhang 1996) as

$$\mathbf{K} = \begin{pmatrix} \frac{f}{s_x} & \frac{f}{s_x}\cot\theta & O_x \\ 0 & \frac{f}{s_y}\sin\theta & O_y \\ 0 & 0 & 1 \end{pmatrix}, \tag{5}$$

[10] Normalized coordinates can be visualized as being unit distance from the optical center, i.e., the focal length is 1.

in which f is the measured camera focal length, s_x the pixel length in the direction of the image scan lines, s_y the pixel length in the direction of the image scan columns, (O_x, O_y) the principal point (intersection of the camera focal or optical axis with its image plane), and θ the skew angle between the pixel axes[11]. With the definition of $\mathbf{K}$ in Equation (5), the relation (Xu & Zhang 1996) between normalized image points and the homogeneous pixel coordinates for the left and right cameras is

$$p_l' = \mathbf{K}_l \bar{p}_l \text{ and } p_r' = \mathbf{K}_r \bar{p}_r. \tag{6}$$

If the expressions from Equation (6) are substituted into Equation (2) and combined with Equation (3), it can be concluded that

$$\mathbf{E} = \mathbf{K}_r^{\mathsf{T}} \mathbf{F} \mathbf{K}_l. \tag{7}$$

Equation (7) provides the relation between the fundamental and essential matrix under the assumption that the cameras are calibrated, i.e., the camera intrinsic parameters are known.

The rank of the essential matrix must equal 2 (Faugeras 1993; Xu & Zhang 1996; Trucco & Verri 1998) and therefore, Equation (7) implies that the fundamental matrix must also have rank equal 2. Since the fundamental matrix is numerically estimated in Step 2, it is unlikely that the estimated fundamental matrix will have rank equal 2.[12] Thus, the essential matrix derived from Equation (7) with the estimated fundamental matrix from Step 2 will generally not have rank equal 2 since both $\mathbf{K}_r$ and $\mathbf{K}_l$ have full rank. However, if the extrinsic parameters (rotation matrix and translation vector) are to be determined in Step 4, the essential matrix must have rank equal 2. In fact, the two non-zero singular values must be equal (Faugeras 1993). These conditions are referred to as rank and singular values constraints. A heuristic approach to enforce these constraints and estimate a matrix "close" to the matrix $\mathbf{E}$ of Equation (7) is to use singular value decomposition (SVD). If the SVD of $\mathbf{E}$ is

$$\mathbf{E} = \mathbf{U} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \mathbf{V}^{\mathsf{T}}, \tag{8}$$

define the modified essential matrix as

---

[11] The angle θ is normally very close to $\pi/2$ and is generally set equal to this value.

[12] Although the assumption that the determinate of the fundamental matrix equals zero is used in the calculations, generally the third singular value is small but not exactly zero because of numerical approximations.

$$\mathbf{E}' = \mathbf{U}\begin{pmatrix} \frac{\sigma_1+\sigma_2}{2} & 0 & 0 \\ 0 & \frac{\sigma_1+\sigma_2}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}\mathbf{V}^{\mathbf{T}}. \tag{9}$$

According to Wang and Tsui (2000), $\mathbf{E}'$ defined in this manner generates a matrix closest to $\mathbf{E}$ in the Frobenius norm satisfying the rank and singular value constraints.

Once $\mathbf{E}'$ is computed, Step 4 is addressed. The following proposition (Faugeras 1993) is germane to the situation.

> ***Proposition***: If a matrix, $\mathbf{E}'$, satisfies the rank and singular value constraints, then it can be factored as the product of a rotation matrix and a skew symmetric matrix containing information about the translation vector, $\mathbf{E}' = \mathbf{R}*\mathbf{S}$.

Specifically, if $\mathbf{T} = \begin{bmatrix} \mathbf{t}_x & \mathbf{t}_y & \mathbf{t}_z \end{bmatrix}^{\mathbf{T}}$ is the translation vector, then

$$\mathbf{S} = \begin{pmatrix} 0 & -\mathbf{t}_z & \mathbf{t}_y \\ \mathbf{t}_z & 0 & -\mathbf{t}_x \\ -\mathbf{t}_y & \mathbf{t}_x & 0 \end{pmatrix}. \tag{10}$$

Since the orientation of the two cameras is unknown, the factorization of $\mathbf{E}'$ is not unique, and as mentioned earlier, the translation vector is known *modulo* a scale factor. The first part of Step 4 determines possible factorizations of $\mathbf{E}'$ while the second part determines the factorization(s) consistent with the assumption that both cameras are pointed in the same direction with depths being positive (i.e., data points are in front of the camera). Determination of the scale factor is discussed in Step 5. According to Wang and Tsui (2000), there are eight different factorizations of $\mathbf{E}'$, consisting of four possible rotation matrices and two possible unit translation vectors. If $\mathbf{U} = (u_1, u_2, u_3)$ and $\mathbf{V} = (v_1, v_2, v_3)$[13] are the two orthogonal matrices from the SVD of $\mathbf{E}'$ in Equation (9), then the eight factorizations are given by $\mathbf{T} = \pm v_3$ and

$$\begin{aligned} \mathbf{R}_1 &= (-u_2, u_1, u_3)\mathbf{V}^{\mathbf{T}} \\ \mathbf{R}_2 &= (-u_2, u_1, -u_3)\mathbf{V}^{\mathbf{T}} \\ \mathbf{R}_3 &= (u_2, -u_1, u_3)\mathbf{V}^{\mathbf{T}} \\ \mathbf{R}_4 &= (u_2, -u_1, -u_3)\mathbf{V}^{\mathbf{T}} \end{aligned} \tag{11}$$

The second part of Step 4, namely, determining the appropriate $\mathbf{R}_i$, $\mathbf{i} = \mathbf{1,2,3,4}$, and the sign of $\mathbf{T}$ consistent with the camera location assumption stated previously, is now addressed.

---

[13] Here, u_i and v_i, $i = 1, 2, 3$, represent the columns of the $\mathbf{U}$ and $\mathbf{V}$ matrices, respectively.

Essentially, the idea is to select a pair of matched image points, compute the sign of the depth or distance from the camera (left point in left camera coordinate system and right point in right camera coordinate system (i.e., $Z_l > 0$ and $Z_r > 0$) for each of the eight combinations of $\mathbf{R}$ and $\mathbf{T}$, and select those $\mathbf{R}$ and $\mathbf{T}$ resulting in both distances being positive.

$\mathbf{T}$ is known *modulo* a scale factor. Therefore, assume that the true value of $\mathbf{T}$ is $\mathbf{T} = \omega\hat{\mathbf{T}}$ in which $\hat{\mathbf{T}} = \pm v_3$. Without the loss of generality, it can be assumed that $\omega > 0$. In the following, $\mathbf{R}$ and $\mathbf{T}$ stand for one of the possible combinations of rotation matrix and translation vector. The objective is to express Z_l and Z_r in terms of the candidate $\mathbf{R}$ and $\mathbf{T}$ and other known quantities. In this case, that includes only the calibration matrices, $\mathbf{K}_l$ and $\mathbf{K}_r$, and matched image points in homogeneous pixel coordinates expressed earlier as p_l' and p_r'.

Following the approach of Trucco and Verri (1998), let ${}_r\mathbf{R}_l$ represent the ith row of the rotation matrix written as a column vector, then from Equation (1)

$$Z_r = {}_r\mathbf{R}_3{}^T(p_l - \omega\hat{\mathbf{T}}). \tag{12}$$

With this expression for Z_r, Equation (4) can be rewritten as

$$\bar{p}_r = \frac{p_r}{{}_r\mathbf{R}_3{}^T(p_l - \omega\hat{\mathbf{T}})}. \tag{13}$$

Next, Equation (1) is used to eliminate p_r in Equation (13) to obtain

$$\bar{p}_r = \frac{\mathbf{R}(p_l - \omega\hat{\mathbf{T}})}{{}_r\mathbf{R}_3{}^T(p_l - \omega\hat{\mathbf{T}})}. \tag{14}$$

Letting $\bar{p}_r = (\bar{x}_r, \bar{y}_r, 1)$, Equation (14) implies

$$\bar{x}_r = \frac{{}_r\mathbf{R}_1{}^T(p_l - \omega\hat{\mathbf{T}})}{{}_r\mathbf{R}_3{}^T(p_l - \omega\hat{\mathbf{T}})}. \tag{15}$$

The next step is to introduce Z_l by replacing p_l with its equivalent expression from Equation (4) to obtain

$$\bar{x}_r = \frac{{}_r\mathbf{R}_1{}^T(Z_l\bar{p}_l - \omega\hat{\mathbf{T}})}{{}_r\mathbf{R}_3{}^T(Z_l\bar{p}_l - \omega\hat{\mathbf{T}})}. \tag{16}$$

Finally, we obtain an expression for Z_l by solving Equation (16) for Z_l.

$$Z_l = \frac{\omega(\bar{x}_{rr}\mathbf{R}_3 - {}_r\mathbf{R}_1)^T\hat{\mathbf{T}}}{(\bar{x}_{rr}\mathbf{R}_3 - {}_r\mathbf{R}_1)^T\bar{p}_l}. \tag{17}$$

Equation (17) is written in terms of normalized image coordinates (i.e., $\bar{x}_r$ and $\bar{p}_l$) instead of the known homogeneous pixel coordinates. Fortunately, Equation (6) provides a way to convert

between the two coordinate systems in terms of the known calibration matrices. In addition, since ω is by assumption positive the final expression for the sign of Z_l is written as

$$S_{Z_l} = \frac{((\mathbf{B}*(\mathbf{K}_r^{-1}\mathbf{p}'_r))_r\mathbf{R}_3 -_r \mathbf{R}_l)^T\hat{\mathbf{T}}}{((\mathbf{B}*(\mathbf{K}_r^{-1}\mathbf{p}'_r))_r\mathbf{R}_3 -_r \mathbf{R}_l)^T\mathbf{K}_l^{-1}\mathbf{p}'_l}. \tag{18}$$

In Equation (18), $\mathbf{B}$ is the row vector $\mathbf{B} = (1,0,0)$.

An expression from which the sign of Z_r can be determined is given in Equation (12), i.e., $Z_r =_r \mathbf{R}_3^T(\mathbf{p}_l - \omega\hat{\mathbf{T}})$. What remains is to express $\mathbf{p}_l$ in terms of known quantities. If the relations in Equations (4) and (6) are combined, $\mathbf{p}_l = Z_l\mathbf{K}_l^{-1}\mathbf{p}'_l$ and from Equations (17) and (18) $Z_l = \omega * S_{Z_l}$. Therefore,

$$Z_r =_r \mathbf{R}_3^T((\omega * S_{Z_l})\mathbf{K}_l^{-1}\mathbf{p}'_l - \omega\hat{\mathbf{T}}), \tag{19}$$

and the sign of Z_r is

$$S_{Z_r} =_r \mathbf{R}_3^T(S_{Z_l}\mathbf{K}_l^{-1}\mathbf{p}'_l - \hat{\mathbf{T}}). \tag{20}$$

Equations (18) and (20) can be used to select the appropriate rotation matrix and translation vector.

Assuming that an appropriate rotation matrix and scaled translation vector are identified, Step 5 consists of determining the scale factor. In order to compute the scale factor, the distance between two distinct points in the scene is required. For UGVs, Haas (2003) suggests the use of two specific locations on the vehicle body, which are easy to identify in an image.

Suppose that a' and b' represent the homogeneous pixel coordinates of the two distinct points in either the left or right camera image (without the loss of generality, assume the left camera image). With the relationship in Equation (6), the pixel coordinates are expressed in normalized image coordinates by

$$\bar{\mathbf{a}} = \mathbf{K}_l^{-1}\mathbf{a}' \text{ and } \bar{\mathbf{b}} = \mathbf{K}_l^{-1}\mathbf{b}'. \tag{21}$$

The normalized image coordinates are converted to 3-D coordinates in the coordinate system of the left camera with the relations expressed in Equations (4), (17), and (18). Thus,

$$\mathbf{a} = \omega * S_{Z_l}\bar{\mathbf{a}} = \omega * S_{Z_l}\mathbf{K}_l^{-1}\mathbf{a}' \text{ and } \mathbf{b} = \omega * S_{Z_l}\bar{\mathbf{b}} = \omega * S_{Z_l}\mathbf{K}_l^{-1}\mathbf{b}', \tag{22}$$

and, since the distance between a and b is known the scale factor, ω can be determined with the distance formula between points in 3-D. If $\|.,.\|$ represents the Euclidean distance between points, the scale factor is expressed as

$$\omega = \frac{1}{|S_{Z_l}|}\frac{\|\mathbf{a},\mathbf{b}\|}{\|\mathbf{K}_l^{-1}\mathbf{a}', \mathbf{K}_l^{-1}\mathbf{b}'\|}. \tag{23}$$

The numerator to the right-hand side of Equation (23) is the known distance between the two locations.

This completes a description of the proposed algorithm. The next section discusses details involving the estimation of the fundamental matrix.

3. Implementation Details Associated With Estimating the Fundamental Matrix

From the description of the proposed algorithm, its success clearly depends on the accuracy of the fundamental matrix estimation in Step 2. However, a number of questions related to the fundamental matrix estimation were left unanswered in Section 2. The specific questions addressed in this section are

1. What error measure should be used in selecting the fundamental matrix?
2. What value for the third homogeneous pixel coordinate, ς, should be used in the calculation of the fundamental matrix?

These questions are addressed empirically with a set of synthetically generated image points based on a given pair of camera calibration matrices and known extrinsic parameters (i.e., rotation matrix and translation vector). To obtain the synthetic image points, 3-D points in the coordinate system of the left camera are randomly generated in the range of $-10 \le X_l \le 10$, $-6 \le Y_l \le 6$, and $1 \le Z_l \le 25$. The limits on the coordinates are chosen so that the pixel coordinates generally fall within a 640x480 grid. Image points that do not fall within this grid are discarded. Distribution of the 500 randomly selected points is shown in Figure 4, indicating that they are fairly uniformly spread and do not provide a degenerative point set. The resulting left image pixel locations are shown in Figure 5 (similar results apply to the right image). Pixel coordinates are computed to sub-pixel accuracy compatible with the MATLAB calculations used to compute the values. Although the image points are not measured, it is still possible that the points contain errors or noise introduced through the calculations. Direct measurement of the error is not possible. However, we can obtain a "sense" for the magnitude of the error in the points combined with the error associated with MATLAB matrix multiplication involving the fundamental matrix by evaluating Equation (2) for the synthetic data points. The fundamental matrix is determined with the same camera calibration matrices, rotation matrix, and translation vector used in the calculation of the points. A histogram of the errors (difference from the true value of 0) obtained by evaluating the right-hand side of Equation (2) for each of the 500 synthetic data points is shown in Figure 6. The average value for the error is 2.16595e-11 and the standard deviation is 1.2945e-10. These values are close to zero, but the possibility of error in the synthetic data points cannot be eliminated. For generating the synthetic data and the calculations,

$$\mathbf{T} = \begin{bmatrix} 0.34768 \\ -0.00637 \\ -0.00608 \end{bmatrix}, \mathbf{R} = \begin{pmatrix} 0.9996879 & -0.006505 & 0.024119 \\ 0.0065039 & 0.999979 & 0.000137 \\ -0.0241202 & 0.000019 & 0.999709 \end{pmatrix},$$

$$\mathbf{K}_l = \begin{pmatrix} 869.314 & 0 & 354.554 \\ 0 & 869.297 & 243.567 \\ 0 & 0 & 1 \end{pmatrix}, \text{and } \mathbf{K}_r = \begin{pmatrix} 839.314 & 0 & 342.382 \\ 0 & 839.245 & 244.141 \\ 0 & 0 & 1 \end{pmatrix}. \quad (24)$$

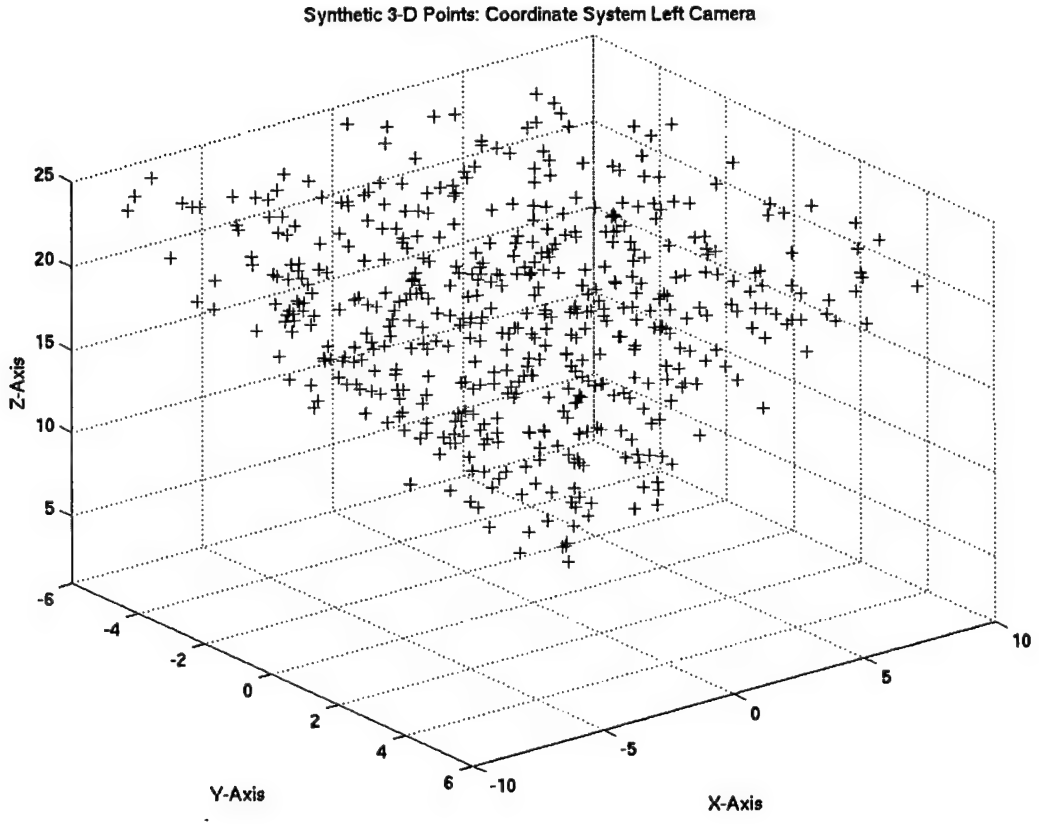


Figure 4. Distribution of synthetic 3-D points in left camera coordinate system.

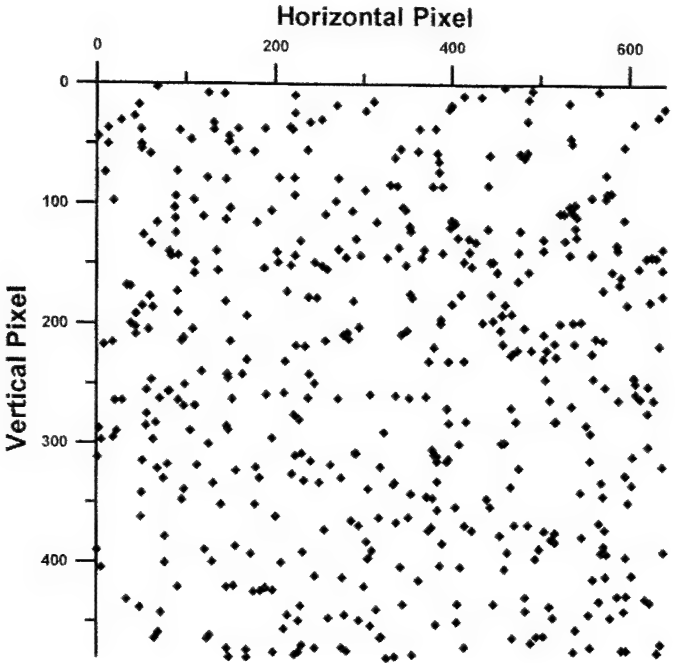


Figure 5. Left image pixel locations of synthetic data.

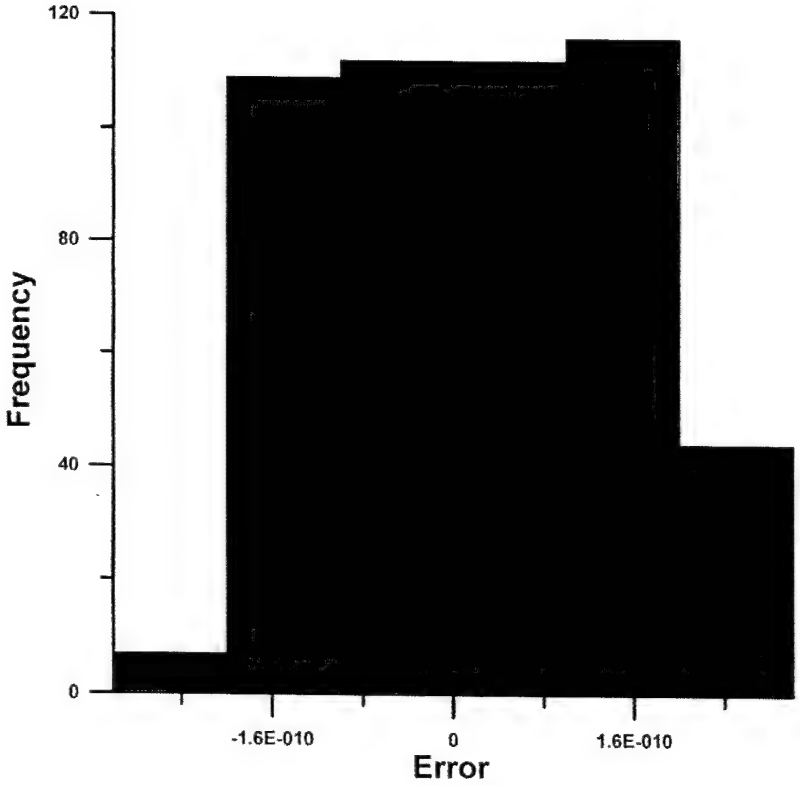


Figure 6. Histogram of errors from evaluation of equation (2) for synthetic data.

Now the question of what error metric to use is addressed. Since the extrinsic parameters are of interest, ideally, an error metric based on these quantities should be used in the selection of the fundamental matrix. However, using a successive approximations approach for the extrinsic parameters causes the same result as the estimation of the fundamental matrix, namely, numerical instability and/or local minimum. Therefore, the error metric chosen is essentially based on the condition expressed in Equation (2), $\mathbf{p}_r'^T \mathbf{F}\mathbf{p}_l' = 0$. Torr (2002) recommends the use of a method adapted from Sampson (1982) and essentially, this is the error matrix recommended for the algorithm. The actual metric used is the sum of the absolute values of the Sampson error computed for each of the points in the data set. Thus, the metric is a single value referred to as the sum of absolute Sampson squared errors (SASSE).

To investigate the effectiveness of the Sampson error metric as being a "good" indicator of the correct fundamental matrix (and thus, the extrinsic parameters), Steps 2 through 4 are coded in MATLAB. Functions from Torr's (2002) structure and motion toolkit are used to perform the MAPSAC, Sampson error metric, and fundamental matrix optimization calculations (a constrained nonlinear estimation enforcing $\det(\mathbf{F})=0$). Outliers are deleted from the matched points between the MAPSAC and optimization calculations.[14] One thousand iterations of Steps 2 through 4 are performed. Each iteration starts with the synthetic data points obtained earlier, so there are 1,000 independent estimates of the extrinsic parameters. Since the true extrinsic parameters, given in Equation (24), are known for the synthetic data points, it is possible to compare Sampson's error metric for the fundamental matrix to error measurements between the true and estimated extrinsic parameters for each iteration. Four different error measurements between the estimated and true extrinsic parameters are used.

If $\mathbf{R}(\text{axis}, \text{rad})$ represents the rotation matrix for a rotation about the indicated axis by rad radians, then the extrinsic parameter $\mathbf{R}$ can be expressed as

$$\mathbf{R} = \mathbf{R}(z,\alpha) * \mathbf{R}(y,\beta) * \mathbf{R}(x,\gamma). \tag{25}$$

In Equation (25), α is the roll angle, β the yaw angle, and γ the pitch angle. The three rotation matrices on the right-hand side of Equation (25) are given by

$$\mathbf{R}(x,\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{pmatrix},$$

$$\mathbf{R}(y,\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix},$$

and

[14] No outliers are computed with the synthetic data, but this procedure is incorporated into the code for use with general data.

$$\mathbf{R}(z,\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{26}$$

Substituting the expressions in Equations (26) into Equation (25) produces the following expression for the rotation matrix in terms of roll, yaw, and pitch angles:

$$\mathbf{R} = \begin{pmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma \end{pmatrix}. \tag{27}$$

Thus, given a rotation matrix, Equation (27) can be used to determine the roll, yaw, and pitch angles. The first three error measures are the difference in milliradians of the roll, yaw, and pitch angles between the estimated and true rotation matrices. The angle in milliradians between the estimated and true translation vectors is the fourth error measure.

With the synthetic data points, a set of extrinsic parameters satisfying the geometric assumptions for the cameras (see Section 2) was determined for 689 of the 1,000 iterations. The minimum SASSE was 5.33E-6. For this iteration, the estimated extrinsic parameters are

$$\mathbf{T}_{est} = \begin{bmatrix} 0.99967895 \\ -0.0183298 \\ -0.0174929 \end{bmatrix} \text{ and } \mathbf{R}_{est} = \begin{pmatrix} 0.9996878 & -0.006506 & 0.024122 \\ 0.0065044 & 0.999979 & 0.000137 \\ -0.0241228 & 0.000020 & 0.999709 \end{pmatrix}. \tag{28}$$

SASSE values and the sometimes rather large deviations from the true extrinsic values are discussed next. A comparison of $\mathbf{T}_{est}$ and $\mathbf{R}_{est}$ with the corresponding values in Equation (24) shows excellent results for the rotation matrix. When we divide $\mathbf{T}_{est}$ component wise by the true translation vector in Equation (24), the vector,

$$\text{ratio} = \begin{pmatrix} 2.87528 \\ 2.87570 \\ 2.87616 \end{pmatrix} \tag{29}$$

is obtained. Thus, to within 0.03% (largest percent difference for the entries of ratio), the estimated and true translation vectors are multiples of each other. As expected with such a close match, the four error measures are extremely close. The largest error is approximately 0.006 milliradian, as shown in Table 1.

Table 1. Errors in milliradians for the estimation of extrinsic parameters with the use of synthetic data and no noise

Error Metric	Error (milliradians)
Yaw	0.00259
Roll	0.00047
Pitch	-0.000134
Translation Vectors	0.00599

Thus, it appears that SASSE is an effective criterion for selecting the fundamental matrix. This is supported by a regression analysis between SASSE and the sum of the absolute values of the four error metrics. Figure 7 shows the graph of the sum of the absolute values of the roll, yaw, and pitch angle errors, combined with the absolute value of the angle between the estimated and true translation vectors versus SASSE for the 689 estimations. The graph clearly shows the wide range of SASSE values, ten orders of magnitude, and total angle error (five orders of magnitude resulting in erroneous estimations of the extrinsic parameters) obtained with the synthetic data. This demonstrates the necessity of performing a number of iterations and not just a single estimation for the fundamental matrix. The regression analysis indicates that SASSE is linearly related to the absolute angle error sum at a confidence level exceeding 99.98%. Thus, it is recommended that the SASSE error metric be used in the selection of the fundamental matrix.

The other implementation question concerned the value to use for the third homogeneous pixel coordinate, ς, required as input to the Torr (2002) routines for computing the fundamental matrix. In the previous calculations, ς is set equal to 1. However, Torr (2002) recommends using a value for ς that provides the best numerical conditioning for his algorithms. Specifically, Torr (2002) suggests that ς be set equal to the estimate of the focal length in pixel units, or if no estimate is available, let $\varsigma = 256$ so that ς is of the same order of magnitude as the image coordinates. Referring to the camera calibration matrices in Equation (24), the estimated camera focal lengths in pixel units are approximately 869 and 839—an average of 854. To investigate the effect of using a different value for ς, the synthetic data calculation from before is re-run for $\varsigma = 256$ and $\varsigma = 854$.

Results comparing the number of times the extrinsic parameters are successfully estimated (of 1,000 attempts) and the minimum SASSE are provided in Table 2. The SASSE for $\varsigma = 256$ and $\varsigma = 854$ are shown in Figure 8. Note that the scales on the vertical axes of the two graphs in Figure 8 differ by an order of magnitude.

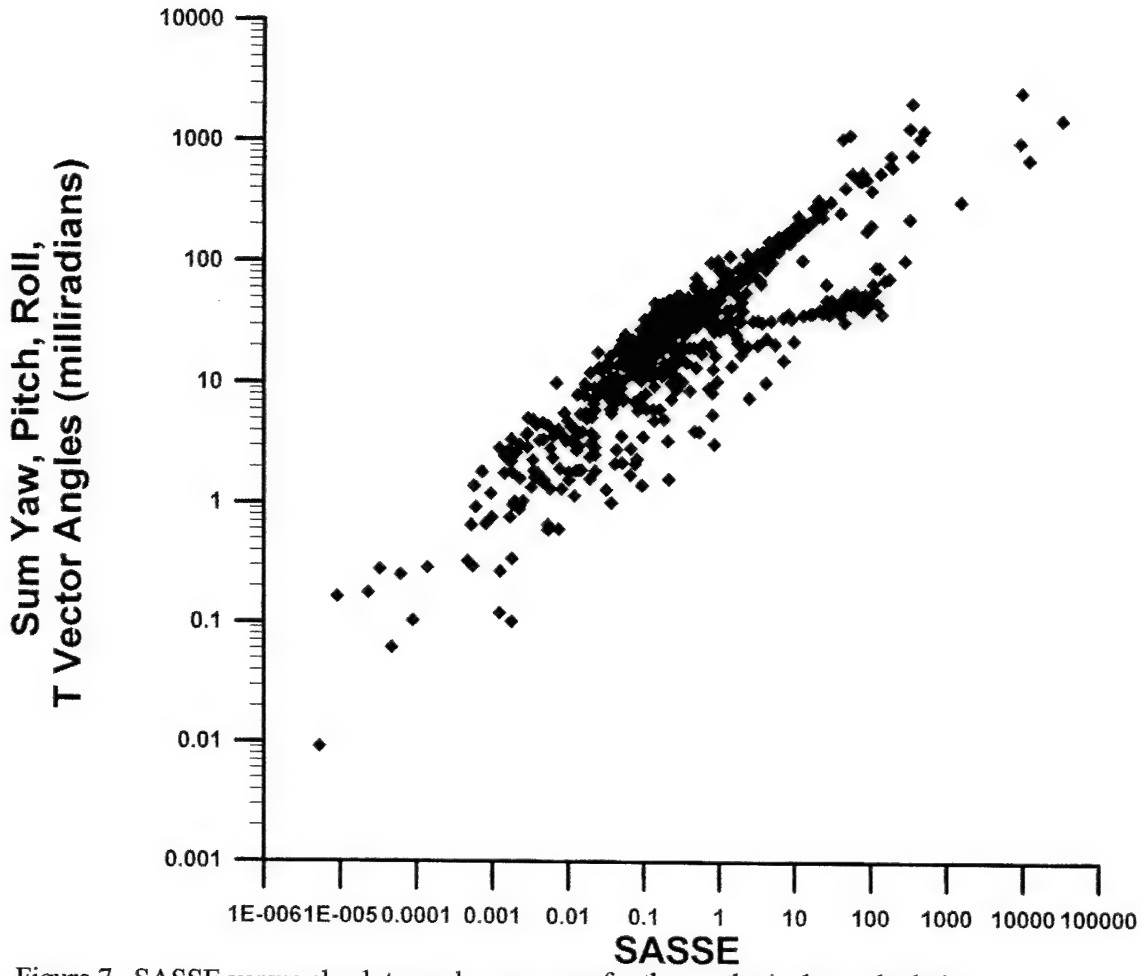


Figure 7. SASSE versus absolute angle error sum for the synthetic data calculations.

Table 2. Results of calculations for different values of ς

ς	**Successful Est. of Extrinsic Parameters**	**Minimum SASSE**
1	689	5.33E-6
256	1000	1.53E-12
854	1000	2.88E-10

From Table 2 and Figure 8, it appears that using the larger values of ς results in more stable numerical calculations (extrinsic parameters estimated in 100% of the iterations versus 68.9% for $\varsigma = 1$) and much lower minimum and overall SASSE values. Unfortunately, a similar trend in the computed extrinsic parameters is not achieved. In fact, the computed rotation matrices and translation vectors are in very poor agreement with the true values, as illustrated in Figures 9 and 10. If one compares Figure 7 with Figures 8 and 9, there is a difference of five orders of magnitude in the minimum absolute angle error sum for $\varsigma = 1$ compared to $\varsigma = 256$ or $\varsigma = 854$. Based on these results, a value of $\varsigma = 1$ is recommended for use in the algorithm.

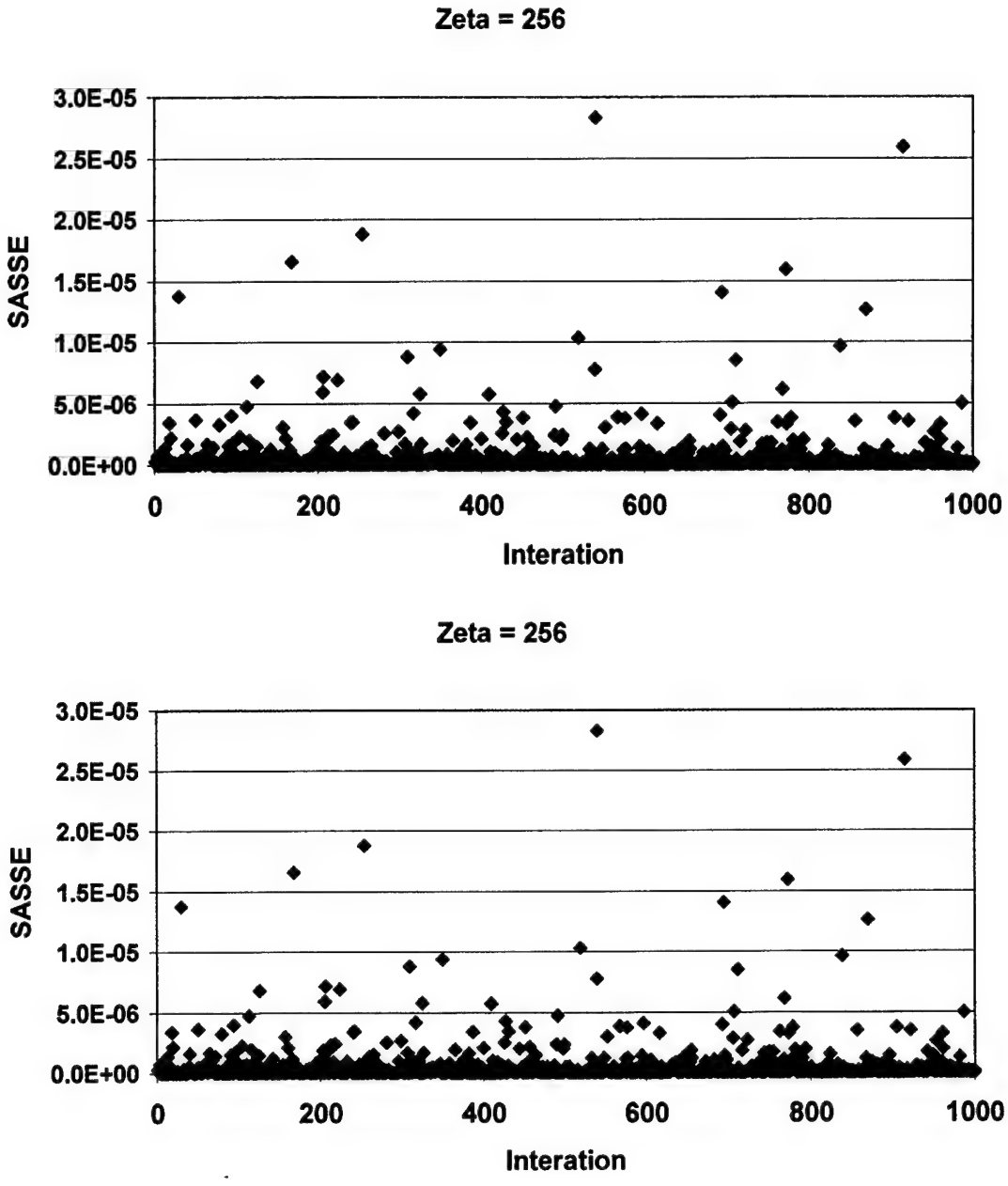


Figure 8. SASSE for ς = 256 (top) and ς = 854 (bottom).

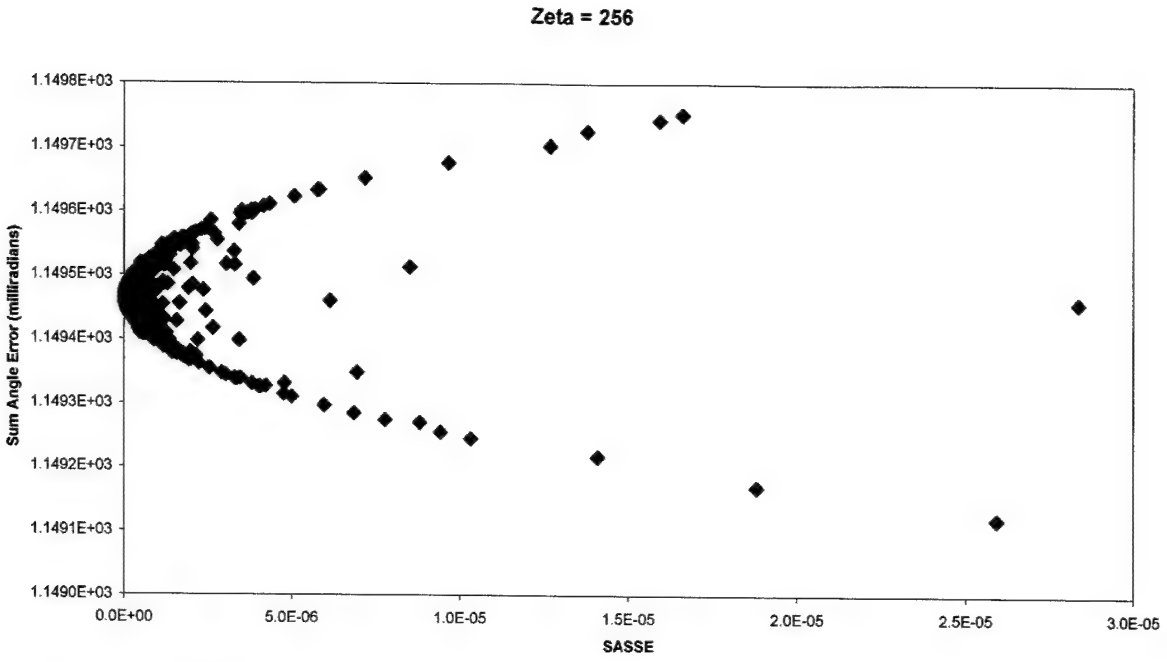


Figure 9. SASSE versus absolute angle error sum, $\varsigma = 256$.

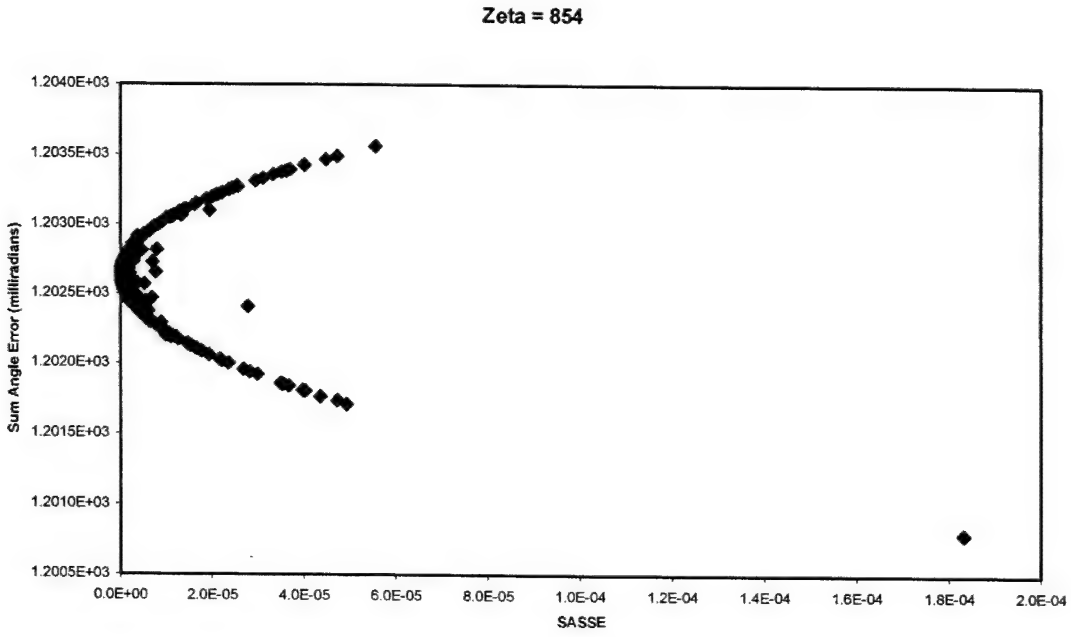


Figure 10. SASSE versus absolute angle error sum, $\varsigma = 854$.

4. Impact of Pixel Location Error

To address the impact of pixel location error on the results of the computed extrinsic parameters, normally distributed random noise with zero mean is added to the synthetic pixel data. The

extrinsic parameters are then computed as in Section 3 for this new set of data. When the variance of the normal distribution used to generate the noise is varied, data sets with different average pixel error[15] are obtained. Results of the calculations are provided in Table 3.

Interpreting the data in Table 3 is difficult since the impact of the error in the extrinsic parameters is meaningful only when the transformation represented by the extrinsic parameters is applied to 3-D points to convert between camera coordinate systems or in image rectification. As expected, the sum of the errors tends to increase as the average pixel error increases, although it is not monotonic. In most cases, the translation vector error dominates the error representing more than 90% of the total error.

Table 3. Errors in extrinsic parameters as a function of average pixel error

Average Pixel Error (Pixel)	Number Inliers (-)	Yaw Error (milliradians)	Roll Error (milliradians)	Pitch Error (milliradians)	Translation Vector Error (milliradians)	Absolute Sum of Errors (milliradians)	Percent Translation Error of Total Error
0.30953	451	-0.30677	0.00075	-0.47925	21.63163	22.67923	95.4
0.31442	445	-1.34523	-0.08148	0.46376	31.33827	33.22876	94.3
0.61760	464	1.21886	0.11332	0.26589	95.01793	96.61600	98.3
0.64231	443	-0.85345	-0.10906	1.19506	134.05170	136.20928	98.4
0.92697	405	-3.04766	1.40773	-5.09897	305.87785	315.43222	97.0
0.93086	440	-1.44756	-0.11436	-1.15306	35.89073	38.60571	93.0
1.11433	397	-1.70679	-1.86785	5.63422	173.82988	183.03873	95.0
1.13192	416	-0.83253	1.29143	-2.10948	319.57389	323.80732	98.7
1.23519	375	313.27911	16.77453	7.76462	191.58728	529.40555	36.2
1.28463	423	63.66804	3.85651	-1.59969	70.28832	139.41256	50.4
1.54495	378	-13.43722	4.73581	-9.48231	343.52339	371.17874	92.5
1.55144	378	-3.06847	-0.84015	4.04731	443.57720	451.53313	98.2
1.87059	370	243.60268	201.13143	33.33670	183.08583	661.15663	27.7
1.92376	325	-4.01233	-2.70836	5.41658	560.02219	572.15946	97.9

These results are somewhat surprising since similar calculations with real data wherein the average pixel error exceeded several pixels generated results closer to the first two lines of Table 3. An analysis of the differences in the data sets showed that the data sets based on real data that provided good estimates for the extrinsic parameters consisted of matched points, most of which corresponded to 3-D points in one of several narrow depth bands in front of the cameras. This is in contrast to the synthetic data that are uniformly distributed in depth.

To determine if restricting the depth of the 3-D data points used to generate the synthetic data improves the extrinsic parameter estimation, a synthetic data set similar to the one used previously is generated with the depths restricted to be in one of three bands: $3 \le Z_l \le 5$, $8 \le Z_l \le 10$, or $13 \le Z_l \le 15$. The data set consists of a total of 501 points equally distributed

[15] Average pixel error is defined as the average distance (Euclidean) in pixels between the original data point and the data point resulting from adding noise. The average is based on both the left and right images. Thus, the total number of distances averaged is 1,000.

among the three depth bands. Results of the extrinsic parameter estimation with the restricted data set with noise added are provided in Table 4.

Table 4. Errors in extrinsic parameters as a function of average pixel error for restricted data points

Average Pixel Error (Pixel)	Number Inliers (-)	Yaw Error (milliradians)	Roll Error (milliradians)	Pitch Error (milliradians)	Translation Vector Error (milliradians)	Absolute Sum of Errors (milliradians)	Percent Translation Error of Total Error
0.30920	478	-0.83248	-0.11850	0.71049	6.38768	8.04915	79.4%
0.31650	478	0.74264	0.19661	0.37298	7.30080	8.61302	84.8%
0.63644	438	1.75895	0.39503	0.06085	22.09261	24.30744	91.8%
0.63702	464	-0.93796	0.10236	-1.67921	30.53573	33.25525	90.8%
0.93900	439	-6.05905	-0.761574	1.18175	22.13129	30.76939	71.9%
0.94186	457	-0.62016	-0.74964	0.96650	15.86392	18.20022	87.2%
1.09381	410	-7.59709	-0.02254	7.56055	9.23383	24.41401	37.8%
1.11380	418	0.47581	0.37349	-4.80996	75.71262	81.37187	93.0%
1.22608	403	-5.84966	0.24586	5.68053	9.88465	21.66070	45.6%
1.22854	416	-9.98535	-0.79442	2.29670	55.10920	68.18568	80.8%
1.56880	377	1.08303	-2.87929	4.35595	154.57800	162.89629	94.9%
1.59592	396	-7.18333	-0.24495	10.82215	53.83216	72.082581	74.7%
1.87672	340	-20.7766	0.13016	5.52011	41.01993	67.44696	60.8%
1.88366	335	-9.21176	-2.84629	5.98874	227.51472	245.56152	92.7%

A comparison of Tables 3 and 4 indicates a substantial decrease in the translation vector error and a more stable estimate of the rotation matrix with the depth restricted data set.

As mentioned before, interpreting the meaning of the errors in Tables 3 and 4 is difficult since the values are presented in isolation. In an attempt to provide a context in which to interpret these errors, the following results are provided. The depth-restricted data set with different amounts of added noise is used to generate a number of estimates of the extrinsic parameters. Yaw, roll, pitch, and translation vector angle errors for two of the estimates representing extremes in the results are given in Table 5. The resulting rotation matrices and translation vectors are then used to perform 3-D reconstruction on the set of 501 matched synthetic pixel points without noise. Three-dimensional reconstruction is performed with the approach discussed in Trucco and Verri (1998), as implemented in Oberle and Haas (2002). We eliminated the unit translation vector scale factor by multiplying the unit translation vector by the length of the translation vector in Equation (24) used in generating the synthetic data points.

Table 5. Yaw, roll, pitch, and translation vector error

Data Set	Yaw Error (milliradians)	Roll Error (milliradians)	Pitch Error (milliradians)	Translation Vector Error (milliradians)
A	-0.13419	-0.08010	-0.08168	9.95826
B	-11.74610	-1.23964	-0.51374	53.22925

For data set A,

$$\mathbf{R}_A = \begin{pmatrix} 0.999692 & -0.006473 & 0.023985 \\ 0.006424 & 0.999979 & 0.000216 \\ -0.023986 & -0.000062 & 0.999712 \end{pmatrix} \text{ and } \mathbf{T}_A = \begin{pmatrix} 0.347719 \\ -0.006673 \\ -0.002632 \end{pmatrix}. \quad (30)$$

Corresponding values for data set B are

$$\mathbf{R}_B = \begin{pmatrix} 0.999910 & -0.005272 & 0.012373 \\ 0.005266 & 0.999986 & 0.000559 \\ -0.123762 & -0.000494 & 0.999923 \end{pmatrix} \text{ and } \mathbf{T}_B = \begin{pmatrix} 0.347484 \\ -0.007842 \\ 0.012369 \end{pmatrix}. \quad (31)$$

The percentage difference in distance between the 3-D reconstructed points and the original 3-D synthetic points on a point-by-point basis is shown in Figure 11 for the reconstruction with $\mathbf{R}_A$ and $\mathbf{T}_A$. The average percent difference in distance for these data is 0.46%. However, the error appears to be clustered into three groups corresponding to the three different depth bands used in generating the synthetic data: points 1 through 167 for $3 \le Z_1 \le 5$ with an average percent error of 0.29%; points 168 through 334 for $8 \le Z_1 \le 10$, average percent error of 0.44%; and points 335 through 501 for $13 \le Z_1 \le 15$ with an average percent error of 0.64%. Thus, not unexpectedly, the error worsens as depth increases. Results for $\mathbf{R}_B$ and $\mathbf{T}_B$ are given in Figure 12.

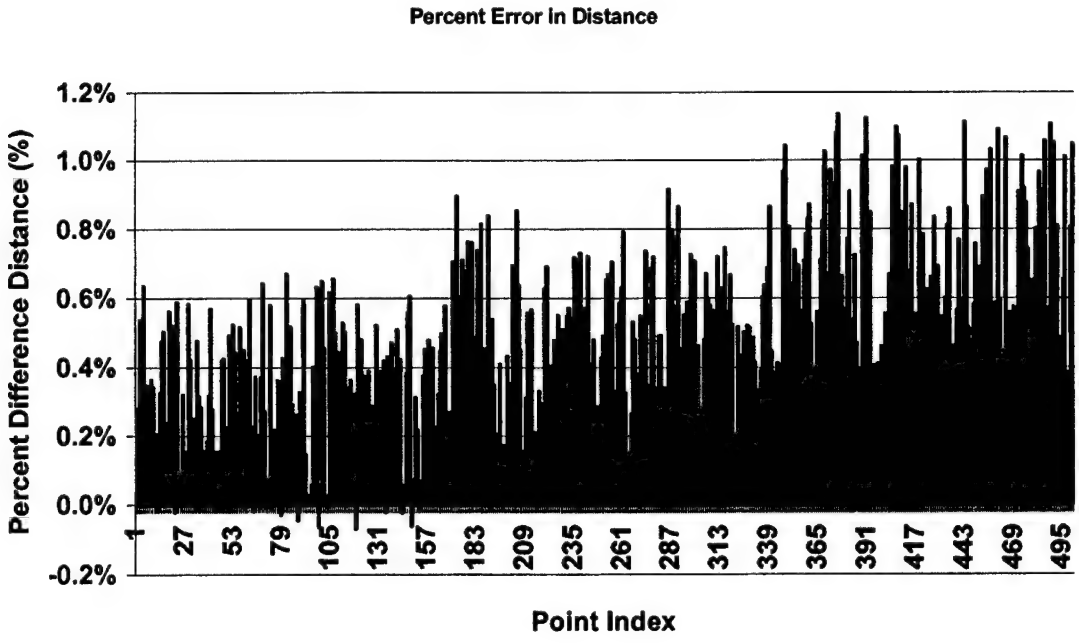


Figure 11. Percent difference in distance between 3-D reconstructed points with extrinsic parameters generated from data set a and synthetic 3-D points.

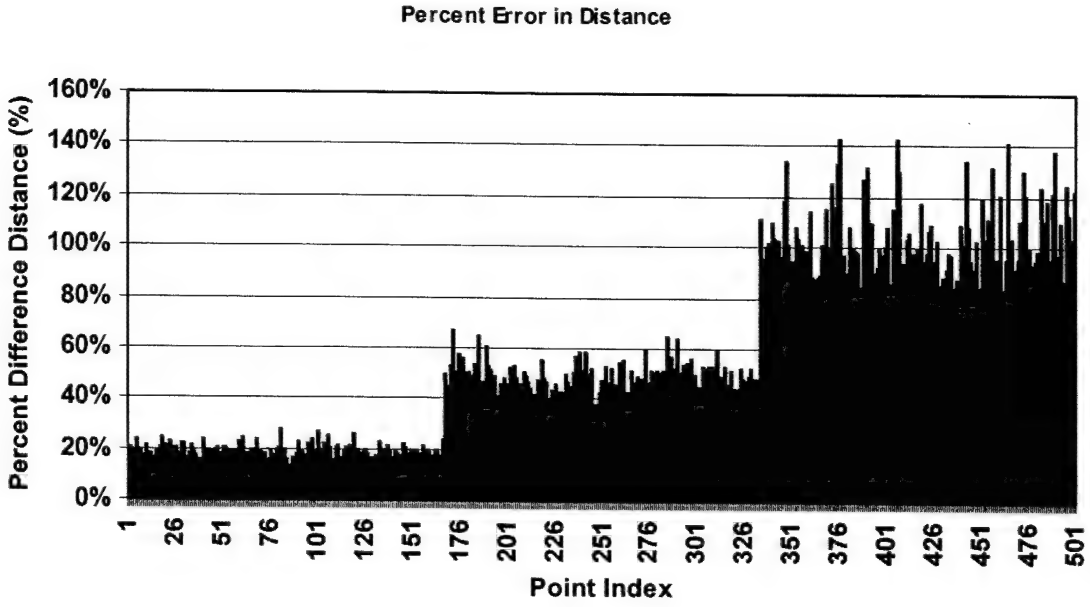


Figure 12. Percent difference in distance between 3-D reconstructed points with extrinsic parameters generated from data set b and synthetic 3-D points.

For data set B, the average percent error has increased by two orders of magnitude to 55.1%. The effect of depth on the error is even more evident than in Figure 11. Looking at Table 5, it is not clear whether the rotation or translation is responsible for the large increase in percent error compared to Figure 11. Although the translation vector error has a much larger magnitude increase (~10 milliradians to ~53 milliradians), the yaw angle error increased by two orders of magnitude while the roll and pitch increased by roughly one order of magnitude, but the maximum magnitude change in the yaw, roll, and pitch angle errors was less than 12 milliradians. To obtain information about the relative importance of the rotation versus the translation in the 3-D reconstruction, two additional 3-D reconstructions are performed:

1. The rotation matrix of data set B is combined with the true translation vector from Equation (24). This calculation addresses the impact of error in the rotation. Results for the percent error in distance are shown in Figure 13.

2. The true rotation matrix from Equation (24) is combined with the translation vector for data set B. This calculation addresses the impact of error in the translation. Results for the percentage error in distance are shown in Figure 14.

Figures 12 and 13 are similar, indicating that the major source of the error in the 3-D reconstruction is attributable to errors in the rotation matrix. A point-by-point comparison between Figures 12 and 13 indicates that the translation vector accounted for only about 3.7% of the total error. This observation is further supported by Figure 14. Not only is the percent error reduced, compared to Figure 12, but no effects appear to be attributable to depth as in Figures 11 through 13.

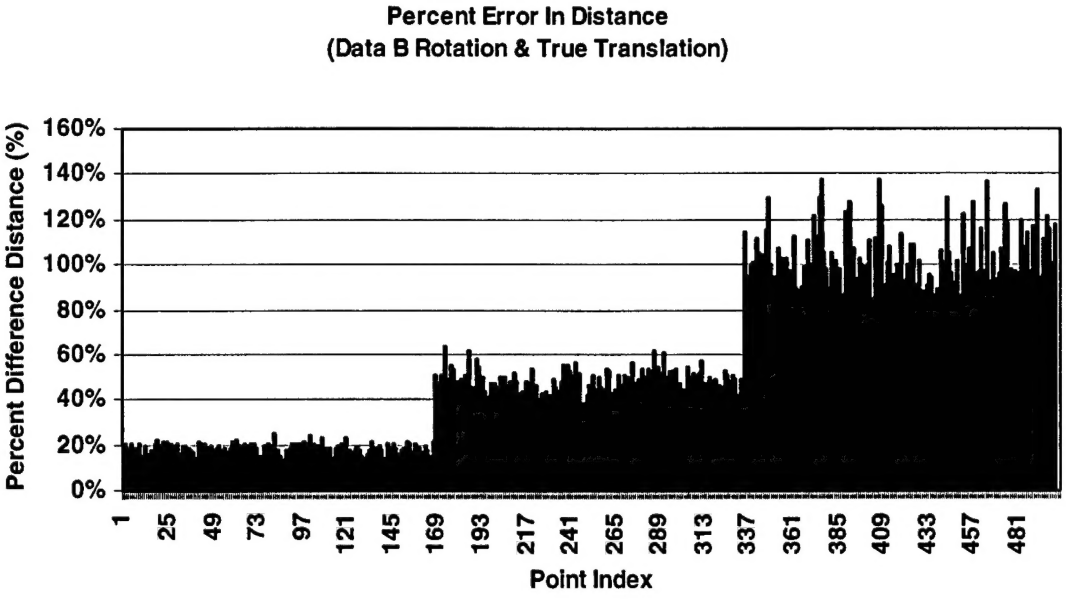


Figure 13. Percent difference in distance between 3-D Reconstructed points with rotation from data set b and true translation vector versus synthetic 3-D points.

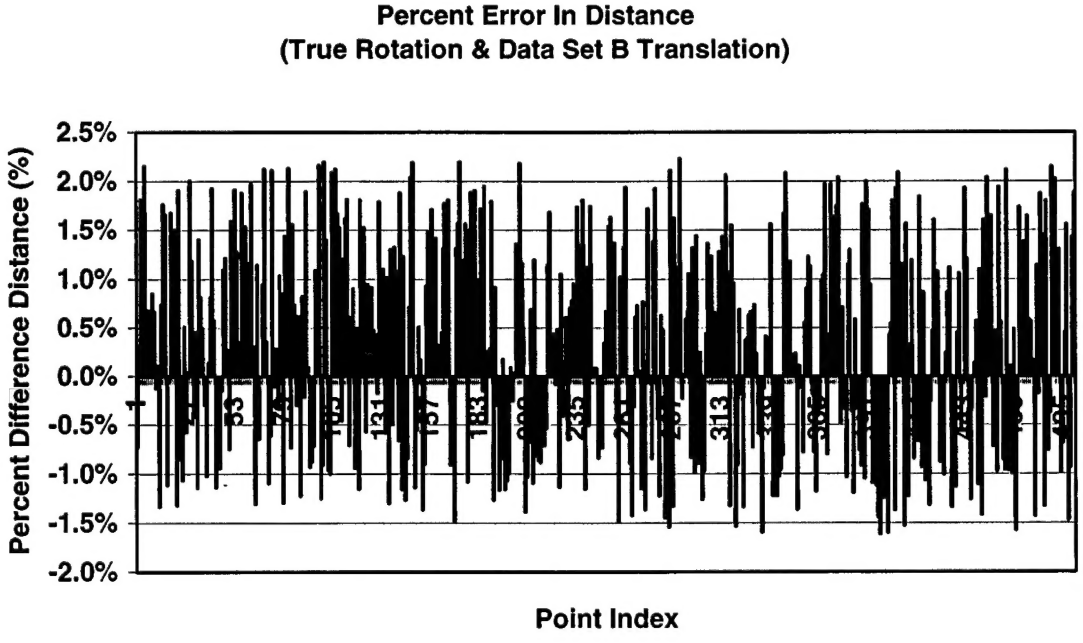


Figure 14. Percent difference in distance between 3-D reconstructed points true rotation and translation vector from data set b versus synthetic 3-D points.

5. Summary

This report explores the potential for performing stereo camera re-calibration to address potential errors in the stereo camera pair's extrinsic parameters. The application of interest is stereo cameras on a UGV, with the errors being introduced because of the UGV's travel over rough terrain.

The algorithm for camera re-calibration presented in Section 2 addresses the two major difficulties associated with this problem, namely, left and right image corresponding point mismatches or outliers and the ill-conditioned fundamental matrix numerical calculation. A RANSAC approach is used to address outliers while multiple iterations of the calculation to minimize the squared errors of the fundamental matrix condition, $p_r'^T \mathbf{F} p_l' = 0$, are employed to overcome the ill-conditioned nature of the calculation. As shown in Section 3, the algorithm appears to provide very good results for noise-free data.

Unfortunately, as shown in Section 4 and well documented in the literature, error in the exact pixel location of corresponding image points can seriously degrade the estimation for the extrinsic parameters. Results from Section 4 indicate that corresponding points should be found with sub-pixel accuracy of 0.5 pixel or less. Restricting corresponding point matches to narrow depth bands also appears to improve the results of the calculation. Finally, it is observed that for the reconstruction problem the major source of error in the 3-D reconstruction is attributable to errors in the rotation matrix with the magnitude of the 3-D reconstruction error highly dependent of the scene depth of the points being reconstructed. The error in the 3-D reconstruction attributable to errors in the translation vector appears to be independent of scene depth.

In summary, it appears that stereo camera re-calibration is possible. However, the accuracy of the re-calibration strongly depends on the ability to determine corresponding left and right image points to sub-pixel precision. Consequently, future work should focus on improving the accuracy of the results of the correspondence problem.

6. References

Faugeras, O. *Three-Dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press: Cambridge, Massachusetts, 1993.

Gennery, D.B. Least Squares Camera Calibration Including Lens Distortion and Automatic Editing of Calibration Points. *Calibration and Orientation of Cameras in Computer Vision*; Gruen and Huang, Eds. Springer Series in Information Sciences, Vol. 34, Springer-Verlag: Berlin and Heidelberg, 2001.

Haas, G.A. U.S. Army Research Laboratory. Private Communication, Aberdeen Proving Ground, Maryland, 2003.

Harris, C.G., Stephens, M. A Combined Corner and Edge Detector, *Proceedings of the 4th Alvey Vision Conference*, pp 147–151, Manchester, United Kingdom, 1988.

Hartley, R.I. In Defense of the Eight-Point Algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **June 1997**, *19(6)*, 580-593.

Lacey, A.J., Pinitkarn, N., Thacker, N.A. An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibration, *Proceedings of the 11th British Machine Vision Conference*, pp 646-655, University of Bristol: United Kingdom, September 2000.

Loy, G. *Calibrated Reconstruction of a Scene*, Lecture Notes for ENGN 4528, The Australian National University, Department of Engineering, FEIT, Canberra, Australia, 2002, web site: www.syseng.anu.edu.au /~gareth/vision_course/online_notes/lecture_3D_ 5_ calibrated_recon.pdf.

Oberle, W.F., Haas, G.A. *Three-Dimensional Stereo Reconstruction and Sensor Registration With Application to the Development of a Multi-Sensor Database*, Army Research Laboratory Technical Report, ARL-TR-2878, U.S. Army Research Laboratory: Aberdeen Proving Ground, Maryland, December 2002.

Sampson, P.D. Fitting Conic Sections to "Very Scattered" Data: An Iterative Refinement of the Bookstein Algorithm, *Computer Vision, Graphics, and Image Processing*, **1982,** *18*, 97-108.

Smith, P., Sinclair, D., Cipolla, R., Wood, K. Effective Corner Matching, *Proceedings of the 9th British Machine Vision Conference*, **September 1998**, 545-556, University of Southampton: United Kingdom.

The MathWorks Inc. MATLAB Version 6.1.0.450, Release 12.1, 2001, web site: www.mathworks.com.

Torr, P.H.S. *A Structure and Motion Toolkit in MATLAB, "Interactive Adventures in S and M,"* Technical Report MSR-TR-2002-56, Microsoft Research: Cambridge, United Kingdom, June 2002, web site: http://research.microsoft.com/~philtorr/.

Trucco, E., Verri, A. *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, Inc.: Upper Saddle River, New Jersey, 1998.

Wang, W., Tsui, H.T. A SVD Decomposition of Essential Matrix with Eight Solutions for the Relative Positions of Two Perspective Cameras, *Proceedings of the 15th International Conference on Pattern Recognition*, Barcelona, Spain **September 2000**, *1*, 362-365.

Xu, G., Zhang, Z. *Epipolar Geometry in Stereo, Motion and Object Recognition A Unified Approach*, Kluwer Academic Publishers: Norwell, Massachusetts, 1996.

Zhang, Z. *Determining the Epipolar Geometry and its Uncertainty: A Review*, INRIA Rapport de Recherche No. 2927, Institut National De Recherche En Informatique Et En Automatique, Sophia-Antipolis, France, July 1996.

Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.T. *A Robust Technique for Matching Two Uncalibrated Images Through Recovery of the Unknown Epipolar Geometry*, INRIA Rapport de Recherche No. 2273, Institut National De Recherche En Informatique Et En Automatique, Sophia-Antipolis, France, May 1994.